DOCUMENT RESUME

ED 227 127 TM 830 110

AUTHOR	Wesson, Caren; And Others
TITLE	Teaching Structure and Student Achievement Effects of Curriculum-Based Measurement: A Causal (Structural) Analysis.
INSTITUTION	Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
REPORT NO	IRLD-RR-105
PUB DATE	Dec 82
CONTRACT	300-80-0622
NOTE	76p.
AVAILABLE FROM	Editor, IRLD, 350 Elliott Hall, 750 East River Road, University of Minneapolis, MN 55455 ($3.00)
PUB TYPE	Reports - Research/Technical (143) -- Guides - Non-Classroom Use (055) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Academic Achievement; Curriculum Development; Elementary Education; Federal Legislation; *Formative Evaluation; *Individualized Education Programs; *Instructional Design; *Instructional Improvement; Program Development; Program Effectiveness; Reading Achievement; Special Education; *Student Evaluation
IDENTIFIERS	Accuracy of Implementation Rating Scale; Minnesota; Structure of Instruction Rating Scale

ABSTRACT

To achieve substantive as well as procedural compliance with Public Law 94-142, it must be determined whether using the formative evaluation system which is useful for monitoring the effects of instruction, increases teacher success in developing student programs. Causal modeling techniques were used to examine the relationships among implementation of a formative evaluation system, structure of instructional programs, and achievement for 117 students in grades 1-7. The Accuracy of Implementation Rating Scale monitored implementation procedures and the Structure of Instruction Rating Scale measured the degree of instructional lesson structure students received. Reading achievement measures were collected three times over the 5-month period by 31 trained teachers. Measurement, structure, and achievement were stable across time and measurement had a short-lived effect on achievement. Measuring student performance had an early effect on achievement, as did silent reading practice. Determining the effect of implementation of an evaluation system or structure of lessons and student achievement was not realized via the present analysis. The appendices contain the Accuracy of Implementation Rating Scale and the Structure of Instruction Rating Scale. (Author/PN)

ED227127

TM

University of Minnesota

Research Report No. 105

TEACHING STRUCTURE AND STUDENT ACHIEVEMENT EFFECTS OF CURRICULUM-BASED MEASUREMENT: A CAUSAL (STRUCTURAL) ANALYSIS

Caren Wesson, Stanley Deno, Phyllis Mirkin, Bonita Sevcik

Russell Skiba, Robert King, Gerald Tindal, and Geoffrey Maruyama

iRLD

Institute for Research on Learning Disabilities

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- [x] This document has been reproduced as received from the person or organization originating it.
- [ ] Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 830 110

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Ysseldyke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Director: James E. Ysseldyke

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with Special Education Programs, Department of Education. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of Special Education Programs.

Research Report No. 105

TEACHING STRUCTURE AND STUDENT ACHIEVEMENT EFFECTS OF CURRICULUM-BASED MEASUREMENT: A CAUSAL (STRUCTURAL) ANALYSIS

Caren Wesson, Stanley Deno, Phyllis Mirkin, Bonita Sevcik,
Russell Skiba, Robert King, Gerald Tindal, and Geoffrey Maruyama
Institute for Research on Learning Disabilities
University of Minnesota

December, 1982

Abstract

Causal modeling techniques were used to examine the relationships among implementation of a formative evaluation system, structure of instructional programs, and achievement for 117 students in grades 1-7. Measures were collected three times over the five-month period. All three constructs were stable across time. Measuring student performance had an early effect on achievement, as did silent reading practice. Limitations of the study and the need for further analyses are discussed.

Teaching Structure and Student Achievement Effects of
Curriculum-Based Measurement: A Causal (Structural) Analysis

In recent years greater demands have been placed on educators, especially special educators, to be accountable for the quality of educational decisions and the ways in which decisions are made. A number of criteria to be followed in assessment and decision-making procedures have been outlined in PL 94-142. Implementation of this part of the law has proved to be difficult due to the absence of technical knowledge that would enable schools to comply with the intent of the law as well as the procedures outlined in the law. In response to this problem, the Institute for Research on Learning Disabilities at the University of Minnesota (IRLD), for the past six years under federal contract, has conducted a program of research and development that has had as its goal developing a functional system for developing and monitoring progress on IEP goals, as intended in PL 94-142.

One objective of this research and development program has been to determine empirically the effects of teachers using the formative evaluation system developed by the IRLD on student achievement in reading, spelling, and written expression. If we are to achieve substantive as well as procedural compliance with the law (Deno & Mirkin, 1982) we must determine whether using the formative evaluation system increases teacher success in developing student programs. In answering this question our focus has been on the IEP adjustment decision that teachers make once special education is being provided for a student. The formative evaluation system is an assessment device for monitoring the effectiveness of the IEP. (See Figure 1.)

The hypothesis is that if an adequate system of formative evaluation is developed, teachers may use this system to monitor student progress and the effectiveness of their instruction. If student progress is not adequate, then teachers judge their instruction to be ineffective and modify their instruction in an attempt to improve the student's progress.

Insert Figure 1 about here

The rationale underlying this hypothesis rests on a set of assumptions. First, the success of special education is defined by the extent to which students' academic and social behaviors are improved. Second, for any mildly or moderately handicapped student, it is impossible to reliably identify special educational alternatives that will be more effective than the regular classroom program. Given the first two assumptions, the initial IEP then must be viewed as a guess about what might be helpful to the student rather than a plan that is guaranteed to help. If the IEP is only a guess, then there is no alternative but to continuously evaluate the effectiveness of the IEP and to modify it when it is unsuccessful. Under such conditions, teachers should be able to increase the success of special education by systematically measuring student progress toward the achievement of program goals and then adjusting student programs to enhance that progress. In a responsive system such as this, student performance data function as the most useful "vital signs" of whether a program is working or should be changed. A evaluation system, when effective,

allows teachers to empirically test their best hunches about how to help students.

One desirable characteristic of a formative evaluation system is that it be useful for monitoring the effects of any type of instruction. For example, whether the teacher chooses DISTAR, a basic sight word method, or any other approach to teach reading, the monitoring system should accurately measure the student's progress in reading, and it must be unbiased with respect to various theoretical approaches to teaching.

Stage One

In order to accomplish the goal of the research and development program, a three-stage plan was designed (Deno, 1979). Stage One included: (a) the identification of the behaviors to be measured in reading, spelling, and written expression, (b) the development of technically adequate measurement procedures for measuring those behaviors, and (c) an exploration of alternative approaches (rule sysems) for using the data generated by these measures to make decisions about the effectiveness of instruction. The studies in Stage One were intended to lay a foundation for subsequent engineering of a generic formative evaluation system. Identifying valid simple measures of student performance was critical since later development of the evaluation system rested on whether performance data that were technically adequate could be easily and frequently collected.

Consistent with the intent of the three-stage plan, measurement and evaluation procedures were developed for three academic areas (reading, spelling, and written expression). The focus of the present

investigation, however, was the use of the procedures when the IEP goal was reading. Therefore, the remainder of this introduction is restricted to reading.

The basic strategy used to identify useful measures involved a process of elimination. Initially, a pool of five easily measured reading behaviors was generated through a review of the available literature. The behaviors measured in reading included: (a) reading isolated word lists; (b) reading isolated words in context; (c) reading aloud from text; (d) identifying deleted words in text; and (e) giving word meanings (Deno, Mirkin, & Chiang, 1982). The next step was to develop simple standardized measurement procedures. Specific directions were devised that could be used routinely to conduct assessment. These specifics included how to choose a sample and provide directions to the student. The third step was to determine the criterion validity of the measurement procedures by correlating the scores obtained from them with scores on commercially available standardized measures, with program placement, and with grade level. The measures that were not reliable or valid, or those that were deemed less acceptable with respect to any other desired characteristics, were eliminated from the pool.

The results of the criterion validity research led to the conclusion that reading aloud from a basal text is an optimal behavior to measure in reading. The rationale for this selection includes the fact that reading aloud provides a broader range of scores than isolated words and relates somewhat more closely to comprehension. In addition, reading aloud requires little teacher preparation since a

teacher can simply randomly select a passage and direct a child to read aloud. The procedures for measuring reading aloud have been detailed elsewhere (Mirkin, Deno, Fuchs, Wesson, Tindal, Marston, & Kuehnle, 1981) and are described in the procedures section of this paper.

Once the procedures had been developed for measuring reading, the next step in Stage One of the research program was to investigate two procedures for writing objectives. Short-term objectives (STOs) are based on the long-range goals, which are developed using a formula and the student's scores from the reading-aloud measure. STOs can be written so that measurement is on a standard task (e.g., reading aloud at a specific level of a reading series) or measurement can be based on a standard criterion applied to sequential tasks (e.g., mastery of units in a basal reader). A survey of teachers who had used both procedures for one school year revealed that most teachers preferred measuring progress in reading through sequential tasks (Fuchs, Wesson, Tindal, Mirkin, & Deno, 1982).

At the same time, several studies were conducted to examine various procedures for using the data generated from the administration of the generic measures. Analyses of student performance data indicated that students showed greater academic growth when a data utilization strategy was in effect than when teachers did not use the data systematically (Martin, 1980; Mirkin & Deno, 1979; Mirkin, Deno, Tindal, & Kuehnle, 1980). Questionnaires designed to evaluate teacher satisfaction with two alternative data-utilization strategies revealed that teachers preferred to use a

combination of the two strategies over using either strategy alone (Fuchs et al., 1982). This finding contributed to the design of the data-utilization strategy employed in Stage Three studies. This strategy is described in the procedures section of this paper.

Stage Two

Stage Two consisted of improving the logistical feasibility (Lovitt, 1977) of the formative evaluation system, as measured by teacher efficiency and satisfaction. No system of formative evaluation would be useful if teachers found it to be too time consuming or if they were dissatisfied with other aspects of the system. Without efficiency and teacher acceptance, the formative evaluation system probably would not be used regardless of its value in monitoring student progress.

A series of field tests was conducted with a cooperating school district. The results indicated that with practice and systematic attempts to reduce measurement time, teachers were able to increase their efficiency 15 times over. At the end of the study teachers required on the average only two minutes to prepare for measurement, conduct a one-minute assessment, and score and graph the results (Fuchs, Wesson, Tindal, Mirkin, & Deno, 1981). These teachers were also highly satisfied with the evaluation procedures. When questioned by independent evaluators the teachers stated that: (a) the system eliminated much of the jargon, ambiguity, and vague descriptions once found in IEPs; (b) the system met the real intent of the law; (c) their own testing was now relevant to the instruction being provided in the classroom; (d) they were confident in the reliability of their

test, making decisions easier and meetings shorter; (e) their testing was more meaningful because a student is compared with peers from his/her own school and grade level; (f) the students were more aware of their own progress because of the frequent charting required by the data-based system; (g) their ability to measure the effectiveness of their teaching strategies with any particular student was improved; and (h) the system made writing IEPs much easier (Wesson, Deno, & Mirkin, 1982). These results clearly suggest that this monitoring system not only is logistically feasible, but, in fact, has practical advantages.

Stage Three

Stage Three of this research and development plan brings the focus of research back to the primary goal: to determine the effects of teachers' use of formative evaluation on student achievement.

This paper is a report of the relationships among the degree of implementation of the formative evaluation system, the amount of structure in the student's instructional program, and the student's rate of academic progress that were obtained during a one-year training and implementation of the formative evaluation system. (See Figure 2.) The hypothesis tested was that the extent to which teachers implement the evaluation system influences the degree to which their teaching is structured, and that structure, in turn, influences the extent to which students demonstrate academic progress. Therefore, the following research questions were addressed:

(1) How well do teachers implement this formative evaluation system given the brief training that was provided?

(2) Is there any relationship between the extent to which the evaluation system was implemented and the degree of structure of the students' instructional programs?

(3) Is there any relationship between the extent to which the evaluation system was implemented and the amount of student achievement?

(4) Is there any relationship between the degree of structure of the students' instructional programs and the amount of student achievement?

Insert Figure 2 about here

Method

Subjects

A total of 31 teachers participated in this study. In this group, there were 26 females and 5 males. On the average, they had 1.9 years of experience teaching regular education and 8.8 years teaching special education. The greatest percentage of teachers (39%) had no experience teaching regular education; 23% had taught special education for one to three years.

There were 117 students included in the study. Their ages ranged from 6 to 13 years, with an average age of 9.5. There were 92 males and 23 females (the sex of two subjects was uncoded) in grades 1-7. The greatest numbers of students were in grades 2-5 (20, 26, 25, and 25, respectively). In grade 1, there were five students, in grade 6, there were nine students, and in grade 7, there were only two students. The students included in the study were, for the most part (111 of the 117), provided with special education in resource rooms.

Measures

Three major types of measures were employed in this study. First, the measure of the degree of implementation of the monitoring system was included since it was critical to know how accurate and complete teachers are in using the evaluation system. Second, measures indicating the degree of structure of the students' instructional programs were included. These measures are useful in determining how the evaluation system influences teaching practices. The third set of measures were student achievement indices. Most of these measures were administered three times during the five-month study.

Implementation variables. The Accuracy of Implementation Rating Scale (AIRS) is an instrument that was developed in conjunction with the manual Procedures to Develop and Monitor Progress on IEP Goals (Mirkin et al., 1981), which was used for teacher training in this study. The AIRS is designed to provide a format by which to monitor the implementation of the procedures described in the manual. The AIRS consists of 12 items rated on a 1 to 5 scale, 1 being the lowest implementation score and 5 being complete and accurate implementation. A complete list of the items and their operational definitions can be found in Appendix A.

Items 1 and 2 of the AIRS, which require direct observation, deal with the accuracy of administration of the measurement and selection of the stimulus materials. Items 3-12 of the AIRS require inspection of various written documents. Specifically, the rater examines the following documents for each student: (a) the Individualized

Educational Plan (IEP), which should specify the long-range goal and short-term objective in reading; (b) the reading graph; (c) the instructional plan for reading; and (d) the record of changes made in the instructional plan in reading. Factors included in items 3-12 pertain to accuracy of establishing: (a) the appropriate measurement level; (b) an adequate baseline, (c) an accurate long-range goal and short-term objective, (d) a detailed graph, (e) a complete instructional program, and (f) the aimline. These items also focus on the timing of instructional changes as well as the types of changes made. (See Appendix A.) The AIRS was used to assess the degree of implementation at the beginning, mid-way, and at the conclusion of this study.

The interjudge agreement for the AIRS ranged from .92 to .98 when percentage of agreement was based on a within one point rating match. The percentage of exact agreement ranged from .73 to .91.

Structure variables. The Structure of Instruction Rating Scale (SIRS) was designed to measure the degree of structure of the instructional lesson that a student received. In this study, the focus was on structure during reading instruction. The variables chosen for inclusion on the SIRS were gathered from current literature on instruction and student academic achievement (cf. Stevens & Rosenshine, 1981). A list of the variables and their operational definitions can be found in Appendix B. Observations were conducted at three different points in time during the study.

The SIRS consists of 12 five-point bipolar rating scales. A rating of 1 is low for the variable and 5 is high. Observers, trained

by videotape to a criterion of .80-.90 inter-rater agreement, rate all variables on the basis of strict definitions at the end of a 20-minute observation period. For the present study, nine research assistants were trained as observers; they reached an inter-rater agreement level of .92 before actually observing in classrooms. The focus of each observation period for the SIRS is on the instructional environment for one student at a time. (See Appendix B.)

The reliability of the SIRS was assessed by means of Coefficient Alpha, a measure of internal consistency. For a sample of 70 students observed in November 1981, the average inter-item correlation was .37, resulting in an alpha of .86. Thus, the SIRS seems to have a high degree of reliability as indexed by a homogeneity measure.

Achievement measures. At three different points in time during the study, three one-minute oral reading measures, consisting of randomly selected passages from the third grade level in Ginn 720, were administered to the students. These measures were selected based on their technical adequacy (Deno et al., 1982) and sensitivity to change (Marston, Lowry, Deno, & Mirkin, 1981). These simple measures are as reliable and valid as traditional standardized tests and yet are more likely to reflect small increments of improvement. The measurements were conducted by directing students to begin reading at the top of the page and continue reading for one minute, at which time the examiner would say stop. If they came to a word they did not know, the examiner would supply the word and prompt them to continue. While the student was reading, the examiner followed along on a copy of the passage and marked errors of substitution and omission.

Following the reading, the numbers of words read correct and incorrect were counted and recorded, with no feedback given to the student. These three reading measures were given at the beginning of the study (pretest), in the middle, and immediately following the final observation (posttest).

Two subtests from the Stanford Diagnostic Reading Test (Karlsen, Madden & Gardner, 1976) also were given as posttest measures. The Structural Analysis and Reading Comprehension subtests were administered along with the final reading passage measures. Each of these subtests has two parts, with Structural Analysis focusing on syllabication (blending and division) and Reading Comprehension focusing on answering both literal and inferential questions for previously read passages.

Procedures

All teachers were trained to carry out a specific set of procedures, including establishing an appropriate measurement level, writing long-range goals (LRGs) and short-term objectives (STOs), collecting three oral reading scores per week for each student, plotting the scores on a graph, and using the data in making decisions about the effectiveness of students' instructional programs.

Measurement. Reading measurement consisted of one-minute timed samples of reading from the student's curriculum. Both words correct and incorrect were scored and charted on equal interval charts. The level of stimulus material for testing, which also became the baseline, was selected as the level from which the student could read aloud between 20-29 words per minute for grades 1 and 2, and 30-39

words per minute for grades 3-6.

Writing goals. Teachers were instructed to write long-range goals for the student's IEP using both the entry level criterion and a desired year-end mastery criterion, usually 70 words correct per minute with no more than 7 errors. The format used in writing the long range goal is shown in Figure 3.

Insert Figure 3 about here

Writing objectives. Two types of short-term objectives were written, performance and mastery; both were based on the long-range goals. For performance objectives, in order to compute the short-term objective, teachers first subtracted the baseline level of performance from the criterion level listed in the LRG. Dividing this difference by the number of weeks necessary until the annual review, they arrived at the number of words per week gain necessary to meet the long-range goal criteria. In performance measurement, the measurement task is a random sample of items from a constant set of stimuli, and the goal is to improve the level of performance on that stimulus material. In graphing performance measurement, the horizontal axis represents successive school days and the vertical axis represents the level of performance on a constant measurement task; each data point represents the level of proficiency on that constant measurement task. The line of best fit through the data points depicts the student's rate of improvement in performance on the set of stimulus material.

When writing mastery based short-term objectives, teachers

backtrack through the reading curriculum to find the level at which the student reads at the mastery rate designated in the long-range goal. The pages or stories between this baseline level and the goal level are counted and divided by the number of weeks until the annual review. This number becomes the criterion used in the STO specifying the average weekly progress necessary to meet the LRG. On the graph, the horizontal axis again represents school days and the vertical axis represents successive segments, pages, or stories of the curriculum mastered. Each data point represents the number of curriculum segments mastered through a given day. The line of best fit through the data points depicts the rate of student progress through the curriculum. The purpose of repeated mastery assessment is to assess the student's rate of mastery in the curriculum, and the purpose of the graph is to display that rate of curriculum mastery. The teacher measures the student on a representative sample of material from the current instructional curriculum unit and plots that level on the graph until mastery is achieved. At that point (a) the teacher registers on the student's graph that a curriculum unit has been mastered, and (b) the set of reading stimulus material on which the teacher measures the student progresses to the next segment in the hierarchy. The two formats used for writing short-term objectives are listed in Figure 4.

Insert Figure 4 about here

Data utilization. In addition to measuring and writing goals and

objectives, the teachers were trained in the use of the measurement procedures for evaluation of the instructional program. In order to monitor student growth, the baseline reading level and the long-range goal were connected by an aimline that showed the students' desired progress. Every seven data points, the teachers were to monitor student growth by means of the split-middle or quarter-intersect method (White & Haring, 1976). An example is given in Figure 5. If the student was progressing at a rate equivalent to or greater than that indicated by the aimline, the instructional program was continued; if the projected rate of growth was less than that indicated by the aimline, teachers were directed to make a major change in the student's instructional program.

Insert Figure 5 about here

Teacher Training

Three formats were used to train teachers in these procedures. For 10 teachers in one special education cooperative, training in the use of the measurement procedures took place in a series of three half-day workshops at the beginning of the school year. Teachers also were provided with the manual, Procedures to Develop and Monitor Progress on IEP Goals (Mirkin et al., 1981), which detailed all the activities teachers were to do. In addition, visits by observers in December, February, and May, and frequent phone contacts, provided feedback to the teachers on the accuracy of their implementation of the measures.

In two other districts, training was conducted by district personnel with the aid of the same manual. In November, three people designated by each district as trainers participated in a one-day trainer's workshop. At this time the procedures were reviewed for the trainers and they were given trainer's manuals that specified activities for them to use when teaching the monitoring procedures to the teachers. After this trainer's workshop, the trainers set up and conducted a series of training sessions in their own districts. Questions about the procedures usually were forwarded to IRLD personnel. On-going phone contact facilitated the training process.

The last type of teacher training involved 10 teachers from a rural special education cooperative that had served as a pilot site. These teachers were trained during one week of full-day workshops prior to the 1980-81 school year and during monthly, half-day workshops throughout the year. These workshops were conducted by IRLD staff and, prior to February, their focus was on training the teachers to (a) write curriculum-based IEPs, (b) create a curriculum-based measurement procedure including mastery and performance systems, (c) measure frequently and graph student progress toward IEP goals, and (d) develop strategies to improve the feasibility of implementing the frequent measurement systems. By February, each teacher had developed curriculum-based IEPs for at least two students and was measuring and graphing those students' reading performance at least three times each week. In February, the data-utilization systems were introduced to the teachers. The remainder of the workshops consisted of teacher presentations of their graphs and discussions of student progress and

changes in instructional plans.

Data Collection

Throughout the year, specific data were compiled by each teacher and sent to an IRLD staff member who was designated as the contact person. Data collection took place on three occasions, separated by approximately two months each, and was synchronized with the SIRS and AIRS observations.

Each teacher compiled a packet for each student in the study consisting of the following forms: (a) SIRS; (b) AIRS; (c) Graph; (d) IEP (IRLD form); (e) Instructional Plan (IRLD form); (f) Changes in Instructional Plan (IRLD form); (g) Student Information Sheet; and (h) 3rd Grade Passage Scores.

To insure confidentiality, each student was assigned an ID number and names were removed before the documents left the district. The information obtained from the teachers was gleaned by research assistants according to the implementation, structure, and achievement variables. On the last round of data collection, teachers were sent the Stanford Diagnostic Reading Tests along with the standard set of forms.

Observer Training

In order to collect SIRS data and rate items 1 and 2 on the AIRS, observations of each student during reading class were necessary. Staff members (lead teachers, program coordinators) from two locations involved in the research carried out the necessary observation procedures in their districts. These observers were trained during one half-day session by two IRLD staff members. A brief review of the

research design was provided at the onset of the training. The primary focus of the training was on actual observation procedures required of the observers throughout the year, particularly proper use of the Structure of Instruction Rating Scale (SIRS) and the Accuracy of Implementation Rating Scale (AIRS).

Explanation of the SIRS included its history and rationale, its purpose, and its administration procedures. Each item on the scale was discussed in detail, including definitions for and examples of several ratings per item. After the SIRS was explained, two videotapes were used as a training aid to give the observers a chance to practice their skills. The tapes consisted of two resource room situations, one demonstrating a model teacher and the other more indicative of a teacher who would receive lower ratings on the SIRS. Each item on both tapes was rated by each observer and an IRLD staff member and discussed. An inter-rater agreement of .80 was required of the observers before the session ended.

The AIRS training consisted of explanations of the two items on the scale that the observers would be rating. The final portion of the training involved the organizational aspect of the data collection. A list of documents that were to be collected at the time of each observation was drawn up and explained. Throughout the year, an IRLD staff member was in contact with the observers on a weekly basis to insure understanding and consistency of the procedures and to answer any questions.

In the other two study sites, trained IRLD staff members conducted the observations. Nine observers were used in one district

and four in the other. Training of these observers was similar to the training of the district personnel. The videotape and code book were presented and ratings were practiced until the interobserver agreement criterion was reached.

Results

The data reported for this study are correlational, limiting any interpretation to statements about the direction of relationships. However, causal modeling techniques provide a method for going beyond descriptions of correlations to making inferences about the logic of directional hypotheses. These findings still cannot be used to prove causality, but, if theoretically justifiable sequencing of variables is possible, can test the plausability of a particular causal model.

The causal modeling analysis is basically a data reduction technique that uses flexible confirmatory factor analysis techniques to display plausible patterns of causal relationships between variables. This approach is called "maximum-likelihood analysis of structural equations" (MLASE). Analysis is facilitated via a computer program, Linear Structural Relations (LISREL), which "simultaneously estimates relations between observed measures and underlying dimensions, relations among the underlying dimensions, and residual variances for dependent underlying dimensions" (Maruyama, Rubin, & Kingsbury, 1981, p. 966).

The constructs of interest in the present causal model analysis were: (a) the implementation of a data-based program modification system for reading performance (Implementation); (b) structure of instructional programs for specific students (Structure); and (c)

reading achievement (Achievement). These constructs are part of the theoretical model displayed in Figure 6 and are described in the methods section under Measures. (See Figure 6.) In this model, which is longitudinal in design, each construct is viewed as being caused by the concurrent constructs and the constructs that temporally preceded it. Within time periods, patterns of influence are hypothesized to go from Implementation to Structure, Implementation to Achievement, and Structure to Achievement. In other words, Structure 3 is a result of Implementation 2 and 3, Structure 2, and Achievement 2. Finally, at Time 3, the scores from the Stanford Diagnostic Reading Test do not cause any other constructs but are caused by all Time 2 and 3 constructs.

Insert Figure 6 about here

In order to analyze data using a causal modeling approach, several methodological limitations must be kept in mind. First, if every variable of interest is included in the causal model, the model will be far too complex to analyze, for there will be more relations than can be specified accurately. Thus, it is important to restrict the model to those variables that the researchers view as crucial. Leaving out some potential contributors to the model may lead to an incorrect interpretation of the phenomenon of interest. Second, only those variables that demonstrate reliability should be included in a causal model since unreliable variables may lead to faulty inferences. Therefore, variables that may be most interesting to researchers must

be left out of the model if they cannot be measured reliably. As a result, important information may not be accessible using a causal modeling approach.

Given these limitations, this approach has two components: (a) the path analysis, which includes all the variables of interest but sacrifices reliability; and (b) the MLASE analyses, which improve reliability but sacrifice some of the critical variables. Each of these components is described later. Prior to these analyses, factor analyses were conducted on the AIRS and SIRS variables in order to establish which variables fit into separate factors.

A factor analysis of the items of the AIRS revealed that 6 of the 12 represented one factor. These six included the items referring to baseline, aimline, instructional level, graph set-up, short-term objective, and long-range goal. These items involve start-up activities that teachers must do in order to begin using the monitoring system; thus, they logically fit into one factor representing measurement. This factor was used in the MLASE analysis. One consequence of this factor analysis was that many of the variables that are crucial to full implementation of data-based program modification were left out of the MLASE analysis. Specifically, the items aimed at assessing implementation of the procedures teachers use to evaluate student progress and then change the student's program (Timing of Instructional Changes, Clear Changes, Substantial Changes) were not included in the analysis because MLASE analyses require reliability of the variables used. Timing of Instructional Changes was added as a construct in the path analysis since it seemed to be

the most critical variable left out of the factor. Because using multiple measures to assess reliability of the variables used is so important for the MLASE analyses, none of these items was used in this analysis.

Factor analysis of the 12 variables on the SIRS revealed that 9 of the 12 defined a single factor that was called Structure. The three variables that did not load on that factor were Independent Practice, Positive Consequences, and Silent Practice on Outcome Behavior. Only the nine variables that defined a single factor were utilized in the MLASE analyses in this study. However, Silent Practice and Positive Consequences were included in the path analysis because additional variables could be added to this analysis and these two variables seemed to be the most critical variables that were not included in the factor.

Path Analysis

The results of the path analysis are shown in Figure 7. The beta weights for the significant paths are given in the figure. Note that the Implementation Construct was renamed as Measurement to highlight the fact that total implementation could not be analyzed in so far as the evaluation items did not load on the factor. The significant paths include the paths from Time 1 to Time 2 and Time 2 to Time 3 for Measurement, Structure, Achievement, and Positive consequences. Other significant paths include Measurement to Structure and Measurement to Achievement ($p < .10$) at Time 2. Also, the paths from Timing of Instructional Changes to Structure and to Positive Consequences at Time 2 were significant. Significant paths at Time 3 include Timing

of Instructional Changes to Structure and to Silent Practice. The path between Silent Practice and Achievement was also significant ($p < .08$). Finally, Achievement 2 and 3 were related significantly to the Stanford Diagnostic Reading Test.

Insert Figure 7 about here

MLASE

As was stated earlier, only the main constructs of the model were used for the MLASE. Those constructs were Measurement, Structure, and Achievement. In order to retain only those factors with demonstrated reliability, several indices of implementation and structure were dropped from the analysis. Since the items dropped were primarily evaluation items, the implementation construct is defined by measurement items. For this reason, that construct was renamed "Measurement." For Measurement, the factor analysis of data on 12 variables revealed that six of these constituted a single factor. The factor loadings ranged from .40 to .93 (see Table 1). These six variables included the following items from the AIRS: (a) Instructional Level (Item 3); (b) Baseline (Item 4); (c) Graph Set-Up (Item 5); (d) Aimline (Item 6); (e) Long-Range Goal (Item 8); and (f) Short-Term Objective (Item 9). Two other variables, Substantial Changes and Clear Changes, loaded on this factor but did not fit conceptually with the other six. The other six variables pertained to the initial set-up of the measurement system whereas the changes variables referred to modification of the instructional program.

Therefore, these two change variables, which also loaded on Factor 3, were not included in Factor 1. At a later time, Baseline also was dropped as a variable because teachers in one of the sites had not been trained to record baseline data on the graphs that were collected for use with the AIRS. This decision was made to avoid significantly lowering the number of cases available for analysis. In sum, the variables that were used as indicators of measurement were Aimline, Graph Set-Up, Instructional Level, Long-Range Goal, and Short-Term Objective. Because MLASE analyses are most effective when there are three or more indicators of each factor, the five variables were randomly assigned to be included in one of three indicators as follows: (a) Instructional Level and Aimline; (b) Graph Set-Up; and (c) Long-Range Goal and Short-Term Objective.

Insert Table 1 about here

A factor analysis also was conducted for the Structure construct. A factor analysis of the 12 items from the SIRS revealed that nine of the variables constituted one factor and three items were not part of this factor (see Table 2). The three excluded items were Independent Practice, Positive Consequences, and Silent Practice on Outcome Behavior. The remaining nine variables were divided randomly into three indicators. The three indicator sets were: (a) Instructional Grouping, Teacher-Directed Learning, and Corrections; (b) Active Academic Responding, Frequency of Correct Answers, and Pacing; and (c) Demonstration/Prompting, Controlled Practice, and Oral Practice on

Outcome Behavior.

Insert Table 2 about here

For the Achievement construct, the factor analysis revealed that words read correctly and errors for the three passages loaded together on one factor (see Table 3). Because the error scores basically mirrored the words read correctly scores, collinearity problems resulted from analyses including both types of scores. Therefore only the words read correctly scores were used as the indicators of the Achievement construct.

Insert Table 3 about here

The next step in preparing for the causal model analysis was to construct a correlation matrix that included the three indicators of each construct (Measurement, Structure, and Achievement) at the three data collection points, plus four end-of-the-year scores from the Stanford Diagnostic Reading Test (Word Blending, Word Division, Literal Comprehension, and Inferential Comprehension). This 31 X 31 matrix was used to estimate reliability and consistency of the indicators of the constructs. The indicators for Structure and Achievement were reliable and stable. The Measurement indicators were less reliable and stable but still considered useful for further analyses.

MLASE analyses were used to estimate the parameter of the model.

The matrix analyzed for the structural equation analysis is a covariance matrix. Because the data are longitudinal, relationships between variables must allow for changes over time in the variance of the variables (e.g., Maruyama et al., 1981). Analysis of standardized correlation matrices would not be appropriate, since they restrict all measures to unit variance and thereby do not allow changes in variability over time.

The model will be explained in two parts, a measurement model and a structural model. The measurement model contains the estimated relations (loadings) between the observed variables and their constructs, the residual variances for observed variables, and the covariances between pairs of residuals for the observed variables. As can be seen from Table 4, all paths were significant as were most of the residual covariances (see Table 4).

Insert Table 4 about here

The structural model contains the estimated relations among the unobserved variables namely, the paths among the constructs of interest. The significant paths that form the structural model are found in Figure 8. (See Figure 8.) All three constructs were very stable across time. Measurement Time 2 is caused by Measurement Time 1, and Measurement Time 3 is caused by Measurement Time 2. Because of the high stability of measurement over time, a couple of paths were dropped from the model since including them caused problems of collinearity. The paths that were dropped include Measurement 2 to

Structure 3 and to Achievement 3. Similar relationships existed for Structure and Achievement. Other significant concurrent causal relationships included Measurement to Achievement at Times 2 and 3. For Time 2, this relationship was positive and for Time 3, it was negative. Also, at Time 3, Measurement was related to Structure, and Achievement 2 and 3 were related to the Stanford Diagnostic Reading Test scores.

Discussion

Many findings are consistent between the path analysis and MLASE analysis. A noteworthy finding is the stability of the three constructs across time. As in the path analysis, the causal model analysis indicates that Achievement, Measurement, and Structure are difficult to impact.

These results are consistent with previous findings that student achievement is very stable over time (e.g., Bloom, 1964, Maruyama et al., 1981; Bloom, 1964; McGarvey, 1978). Maruyama et al., (1981) examined the relationships among achievement, self-esteem, social class, and ability, using a sample of 715 children aged 9-15. They found that achievement was very stable. They noted that "not even a variable such as ability seems to exert any incremental influence on achievement" (p. 972). The students in the present study fell within the same age range and stability of achievement was equally as evident. This finding was discouraging, as our hope was to make some impact on achievement.

Measurement is also stable, which indicates that teachers who initially learn to implement the measurement procedure accurately

continue to do so, while for teachers who are initially less skillful in measurement, practice is not sufficient to improve their measurement skills. Clearly, teachers who do not implement measurement procedures to criterion as a result of initial training must be targeted for more intensive training if full implementation of the system is desirable.

The Structure construct displays the same degree of stability as that obtained for Achievement and Measurement. Apparently, if a teacher designs a highly structured program for a student, that student continues to receive highly structured instruction throughout the school year. In contrast, students whose instruction is less structured initially also continue to receive less structured education throughout the year. Since the hypothesis contained in the causal model is that implementation of the evaluation system will increase structure, the hypothesis is not supported by the findings. However, the failure to measure the evaluation and change components of the system renders the test of this hypothesis inadequate.

Also common to both analyses is the relationship between Measurement and Achievement at the middle of the study. As others have shown (Jenkins, Mayhall, Peshka, & Townsend, 1974), measuring student performance can result in increased performance. Thus, while measurement alone is not intended as a sufficient condition for affecting student performance in the model, measurement alone does seem to operate directly on achievement. Since the relationship between Achievement as measured by reading aloud from a basal text and Achievement as measured by the Stanford Diagnostic Reading Test is

strong and positive, the operation of measurement alone must be considered a dependent variable that will affect reading proficiency in general. The strong positive relationship between reading aloud and general reading achievement is consistent with past results (Deno et al., 1982), which established the validity of reading aloud as a measure of reading proficiency.

Several important findings deserve to be highlighted. First, during the study Measurement had a strong effect on two other major constructs, Structure and Achievement. This was the expected relationship given the rationale for the data-based program modification procedures. Troublesome, however, was the fact that these effects seem to be shortlived and were not manifest at the end of the study as hypothesized. Perhaps measurement has a short term positive effect on Structure and Achievement, but that as reactivity to measurement decreases, more sophisticated procedures (such as evaluation of student performance data) and adjustments in the instructional program need to be implemented if the potential benefits of measurement are to be realized.

This hypothesis receives some support from the beta weight reported for Timing of Instructional Changes in the path analysis. In the middle of the treatment period the extent to which teachers properly timed instructional changes (as indicated by the data) was negatively related to Achievement. Thus, perhaps measurement activities are important early on in the implementation of data-based program modification, but the positive effects of measurement cannot be sustained unless evaluation procedures also are used.

The lack of an impact of Structure on Achievement is troublesome for the causal model. In the model, Structure is hypothesized to directly influence achievement. This lack of relationship may be because the SIRS does not validly measure the structure variables affecting achievement that others have identified (Stevens & Rosenshine, 1981). A more likely reason may be methodological. Although, the SIRS has established validity (Deno, King, Skiba, Sevcik, & Wesson, 1982), the sampling procedure used in this study weakens its utility for longitudinal research. Data collected on structure on three occasions for a total of 45-60 minutes of instruction over a five to seven month period may not be a good representation of the structure of instruction the student received on a daily basis over that time period. Evidence of sampling bias is suggested in comments made by teachers, who indicated that instruction looked different on days observers were present.

Of special interest is the relationship between silent practice and achievement found in the path analysis. This relationship has been obtained previously by Leinhardt, Zigmond, and Cooley (1980) and Thurlow, Graden, Greener, and Ysseldyke (1982). The consistency of this finding across researchers provides a firm empirical base for the proposition that silent reading practice is an activity that significantly improves general reading proficiency. Sufficient evidence has been amassed to recommend to teachers that they plan for, and provide, increased amounts of silent reading practice for students as a part of their daily reading program. Such a recommendation takes on increased importance when considered in light of the relatively

small proportion of time actually allocated to silent reading by teachers, and the small amount of time students actually engage in silent reading (Leinhardt et al., 1980; Graden et al., 1982; Thurlow et al., 1982).

With regard to the actual ratings received on the AIRS and SIRS, there are several interesting points. Basically, teachers were adequately trained to conduct measurement activities and to write goals and objectives. The mean ratings for most of the AIRS items were above 3.5 (5 being complete and accurate implementation). The items on which teachers received the lowest scores were Instructional Level, Timing of Instructional Changes, and Substantial and Clear Changes. Basically, it appears that teachers were less successful in mastering the parts of this system aimed at evaluating student progress. Not only were the mean ratings lower on these items, but many teachers made no instructional changes for many students. That is, the majority of students were instructed with their original plan for the entire duration of the study.

Mean scores on the SIRS were more variable, ranging from 1.71 to 4.36. The highest mean scores were for items concerning Teacher-directed Learning, Active Academic Responding, Frequency of Correct Answers, and Correction Procedures. The lowest mean scores were on Positive Consequences, Independent Practice, and Silent Reading. Teachers used few positive reinforcers, other than praise, and seldom provided feedback during independent practice. Also, relatively little time was dedicated to silent reading.

While many of the present findings are interesting, several

hypotheses were not tested. First, due to the early factor analysis conducted on the measure constructed to scale implementation, several important variables were dropped from this analysis. Thus, while the analysis tested the hypothesis that measuring student performance and goal-setting affect structure and achievement, it was not possible to determine whether data utilization components would affect structure and achievement. In the general data-based program modification model, student performance data are to be charted and used to evaluate the effectiveness of instruction. If instruction is insufficient for goal attainment, a change is to be made in the student's instructional program. In such a system, measurement is viewed as necessary but not sufficient to effect optimal student growth. That evaluation procedures will affect Structure and Achievement was supported by the path analysis finding that the timing of instructional changes affected Silent Practice, which in turn affected achievement. In addition, it appears that Measurement may impact Structure. In the causal model, the impact of Measurement on Structure at Time 3 is consistent with model hypotheses. The routine of measuring student progress over time apparently results in teachers increasing the structure of their lessons. If this is the case, then if the evaluation components were implemented completely they would probably yield an even stronger causal effect of the continuous evaluation system.

To summarize, the main conclusions of this causal analysis are that measurement, structure, and achievement are stable across time and that measurement has a short-lived effect on achievement. In

addition, silent practice in reading seems to relate to achievement gains. Finally, the primary goal of determining the effect of the implementation of an evaluation system on structure of lessons and student achievement was not realized via the present analysis. Hopefully, further analyses will achieve this goal.

References

Bloom, B. S. Stability and change in human characteristics. New York: Wiley, 1964.

Deno, S. L. Behavioral research methodology as a basis for the formative evaluation of learning disability services. In J. E. Ysseldyke & P. K. Mirkin (Eds.), Proceedings of the Minnesota Roundtable Conference on Assessment of Learning Disabled Children (Monograph No. 8). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Deno, S., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance. Teaching Exceptional Children, 1982, 49, 36-45.

Deno, S., King, R., Skiba, R., Sevcik, B., & Wesson, C. A scale for assessing the structure of instruction in resource classrooms (Research Report, in preparation). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Deno, S., Mirkin, P., & Chiang, B. Identifying valid measures of reading. Exceptional Children, 1982, 49(1), 36-45.

Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. L. Teacher efficiency in continuous evaluation of IEP goals (Research Report No. 53). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981.

Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Graden, J. L., Thurlow, M. L., Ysseldyke, J. E., & Algozzine, B. Instructional ecology and academic responding for students in different reading groups (Research Report No. 79). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Jenkins, J., Mayhall, W., Peshka, C., & Townsend, V. Using direct and daily measures to increase learning. Journal of Learning Disabilities, 1974, 10, 604-608.

Karlsen, B., Madden, R., & Gardner, E. F. Stanford diagnostic reading test. New York: Harcourt Brace Jovanovich, 1976.

Leinhardt, G., Zigmond, N., & Cooley, W. Reading instruction and its effects. Paper presented at the annual meeting of the American Educational Research Association, Boston, 1980.

Lovitt, T. In spite of my resistance, I've learned from children. Columbus, OH: Charles E. Merrill, 1977.

Marston, D., Lowry, L., Deno, S., & Mirkin, P. An analysis of learning trends in simple measures of reading, spelling, and written expression (Research Report No. 49). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981

Martin, M. D. A comparison of variations in data utilization procedures on the reading performance of mildly handicapped students. Unpublished doctoral dissertation, University of Washington, 1980.

Maruyama, G., Rubin, R., & Kingsbury, G. Self-esteem and educational achievement: Independent constructs with a common cause? Journal of Personality and Social Psychology, 1981, 40, 962-975.

McGarvey, W. E. Longitudinal factors in school desegregation (Doctoral dissertation, University of Southern California, 1978). Dissertation Abstracts International, 1978, 38, 5097B.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979. (ERIC Document Reproduction Service No. ED 185 754)

Mirkin, P. K., Deno, S. L., Fuchs, L., Wesson, C., Tindal, G., Marston, D., & Kuehnle, K. Procedures to develop and monitor progress on IEP goals. Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1980. (ERIC Document Reproduction Service No. ED 197 510)

Stevens, R., & Rosenshine, B. Advances in research on teaching. Exceptional Education Quarterly, 1981, 2(1), 1-10.

Thurlow, M. L., Graden, J., Greener, J. W., & Ysseldyke, J. E. Academic responding time for LD and non-LD students (Research Report No. 72). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Wesson, C., Deno, S., & Mirkin, P. Research on developing and monitoring progress on IEP goals: Current findings and implications for practice (Monograph No. 18). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

White, O., & Haring, N. Exceptional teaching: A multi-media training package. Columbus, OH: Charles E. Merrill, 1976.

Footnote

The authors gratefully acknowledge the participating school districts for their cooperation in this research, and the assistance of Jean Greener in coordination of data collection.

Table 1

Factor Loadings for the Measurement Variables

	Factor 1	Factor 2	Factor 3
Administration	-.04	.36	.79[3]
Selecting Stimulus Material	.06	.63[2]	.28
Instructional Level	.69[1]	.18	-.24
Baseline	.93[1]	.07	.02
Graph Set-Up	.40[1]	.04	.34
Aimline	.58[1]	-.13	-.04
Timing of Instructional Changes	.08	.94[2]	-.32
Long-Range Goal	.47[1]	.31	.31
Short-Term Objectives	.44[1]	-.19	.34
Instructional Plan	.07	-.13	.37[3]
Substantial Changes	.75	.12	.43[3]
Clear Changes	.58	.16	.38[3]

[1]
[2]Items which load on Factor 1.
[3]Items which load on Factor 2.
Items which load on Factor 3.

Note: Substantial Changes and Clear Changes were seen as Factor 3 since Factor 1 included items pertaining to the set up of the measurement system; both change items are pertinent to using the data in an evaluative manner.

Table 2

Factor Loadings for the Structure Variables

	Factor 1
Instructional Grouping	.45[1]
Teacher-Directed Learning	.66[1]
Active Academic Responding	.59[1]
Demonstration/Prompting	.66[1]
Controlled Practice	.70[1]
Frequency of Correct Answers	.36[1]
Independent Practice	.20
Corrections	.55[1]
Positive Consequences	.16
Pacing	.64[1]
Oral Practice on Outcome Behavior	.52[1]
Silent Practice on Outcome Behavior	-.12

[1]Variables which load on Factor 1.

Table 3

Factor Loadings for the Achievement Variables

	Factor 1
Passage 1 - Correct	.79[1]
Passage 2 - Correct	.78[1]
Passage 3 - Correct	.75[1]
Passage 1 - Error	-.75[1]
Passage 2 - Error	-.76[1]
Passage 3 - Error	-.73[1]

[1]Variables loading on Factor 1.

Table 4

Relationships Among Variables Shown in Figure 6

Measurement Model					
Loadings		Residuals		Covariance of Residuals	
A	.79	a	.65	ai	.353*
B	.77	b	.83	bj	.415*
C	.60	c	.13	br	.503*
D	.91	d	.13	cr	.124
E	.87	e	.26	cs	.175*
F	.84	f	.34	ia	.280*
G	.95	g	.04	jr	.391*
H	.93	h	.15	ke	.288*
I	.78	i	.40	yz	.087
J	.73	j	.53	aabb	.194*
K	.71	k	.39	em	.080*
L	.81	l	.20	aq	.329*
M	.70	m	.26		
N	.79	n	.56		
O	.92	o	.07		
P	.90	p	.22		
Q	.68	q	.36		
R	.60	r	.86		
S	.60	s	.39		
T	.77	t	.10		
U	.74	u	.38		
v	1.00	v	.40		
W	.95	w	.13		
X	.81	x	1.16		
Y	.90	y	.55		
Z	.74	z	.32		
AA	.86	aa	.42		
BB	.77	bb	.26		

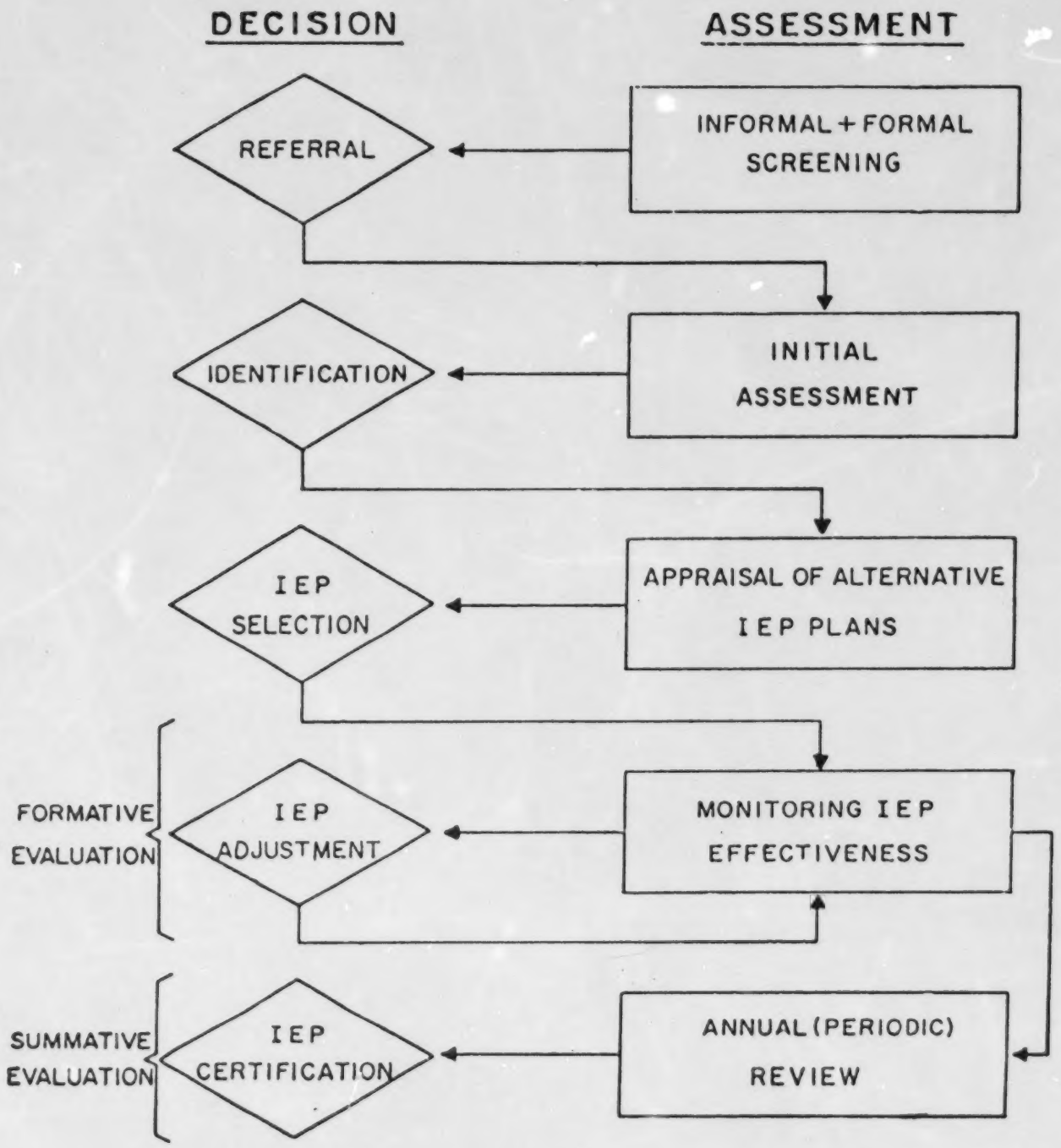

Figure 1. Special Education Decision Making Processes

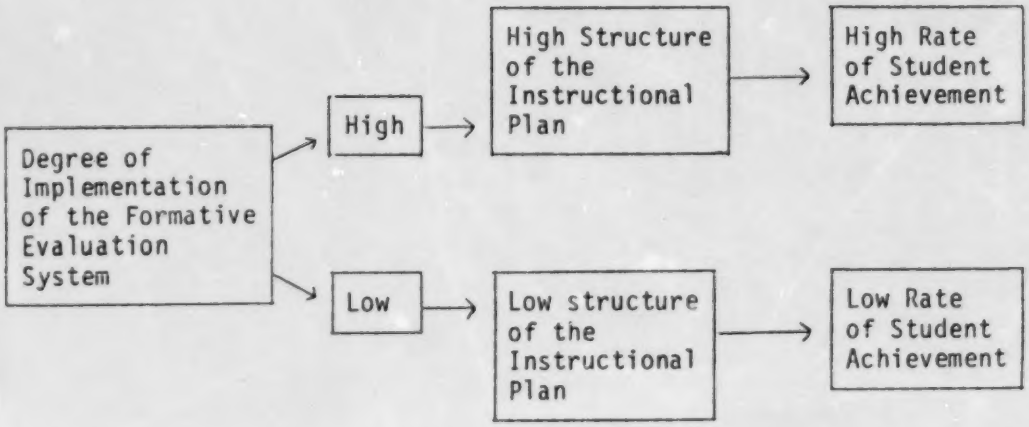

Figure 2. Causal Model Research Design Hypothesis

	Condition	Behavior	Criteria
LRG:	In _____ (total # weeks) weeks, when presented with stories from Level _____ (#), _____ (reading series),	student will read aloud	at the rate of 50 wpm or better 5 or fewer errors.

Figure 3. Format for Long-Range Goal: Reading

Performance Charting: Reading

CONDITION	BEHAVIOR	CRITERIA
Each successive week, when presented with a random selection from_____ (level # from current instructional level - same as LRG) of_____ (reading series)	student will read aloud	at an average increase of _____ (70 or 50 wpm - actual performance) total # weeks remaining in school year.

Progress Charting: Reading

CONDITION	BEHAVIOR	CRITERIA
Each week, when presented with successive stories from _____ (Level #s from current instructional level to annual goal level)	student will progress	at the rate of _____ stories per week maintaining the mastery criteria of at least 50 wpm (gr. 1 & 2) with 5 or fewer errors and 70 wpm (gr. 3-6) with 7 or fewer errors

Figure 4. Performance and Progress Charting Short Term Objectives for Reading.

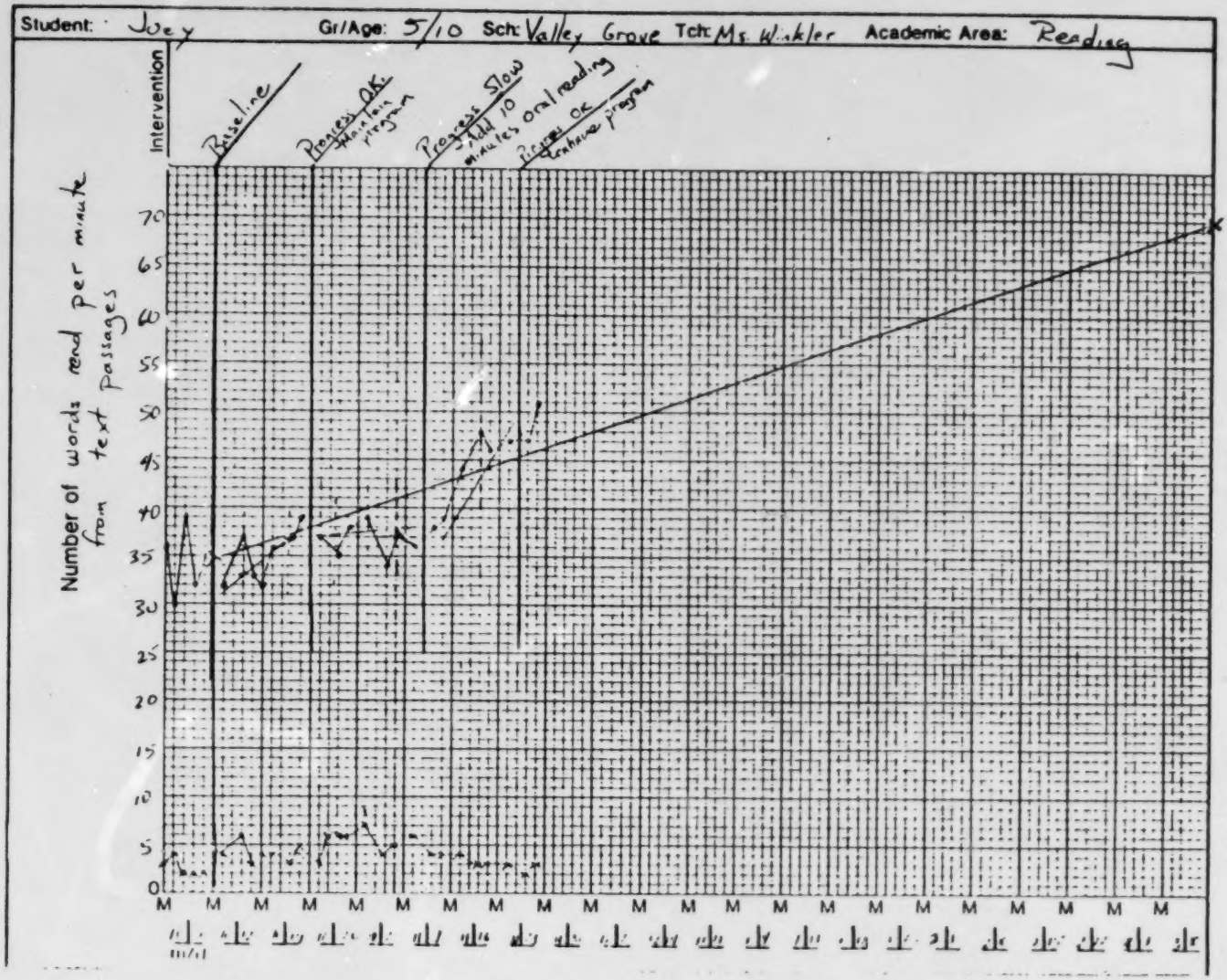

Figure 5. Using the split-middle technique to monitor student progress.

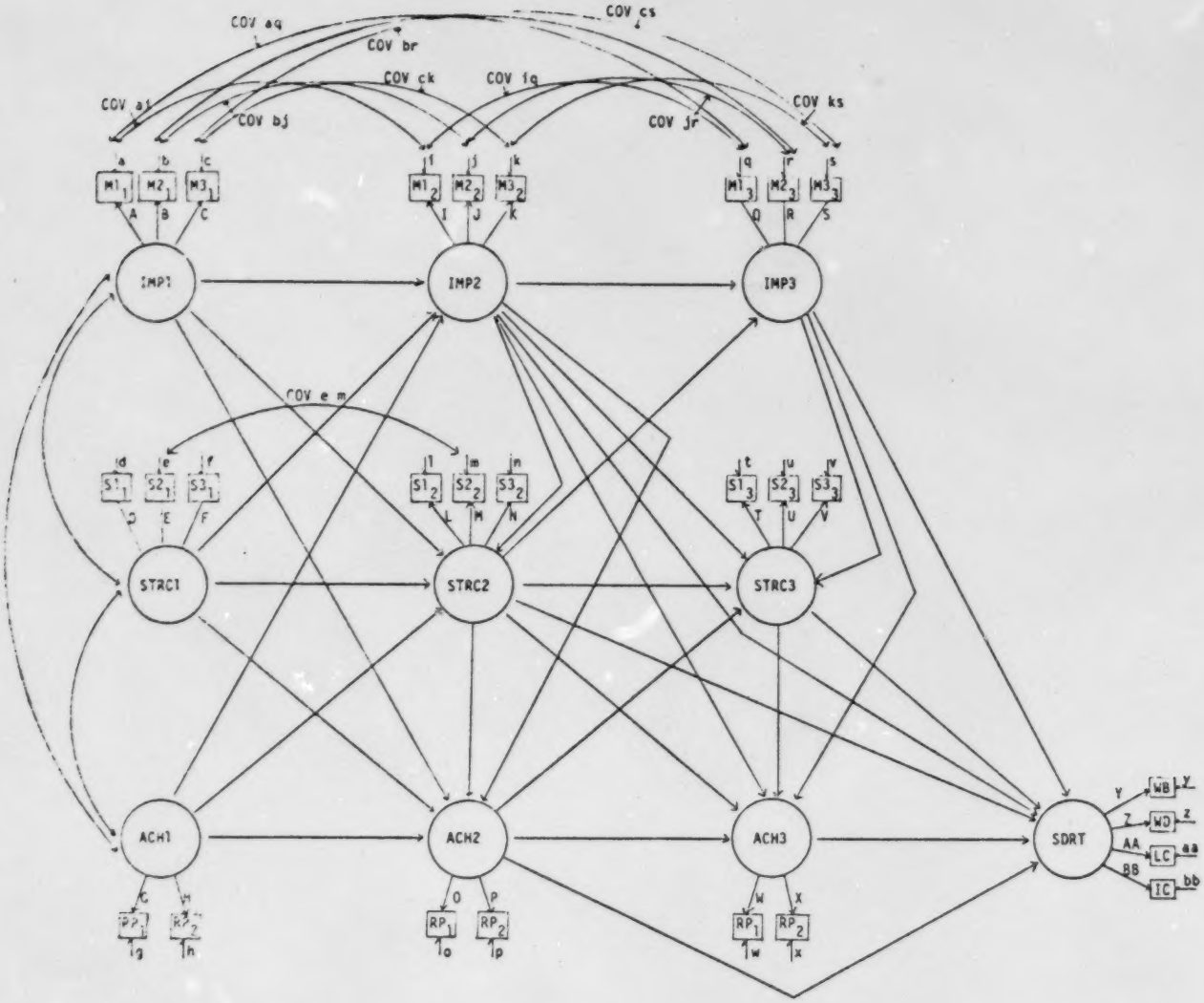

Figure 6. Theoretical Model depicting interrelationships among implementation (IMP), structure of instruction (STRC), achievement (ACH), and Stanford Diagnostic Reading Test (SDRT). The measures are measurement indicator 1,2, and 3 (M1, M2, M3), structure indicators 1, 2, and 3 (S1, S2, S3), reading passages 1 and 2 (RP1, RP2), and subtests from the Stanford Diagnostic Reading Test, Word Blending (WB), Word Division (WD), Literal Comprehension (LC) and Inferential Comprehension (IC). The subnumeral represents the time when each datôm was collected. Observed measures are represented by rectangles and the constructs underlying the measures are represented by circles. Causal paths are illustrated by straight arrows while relationships in which causal relationships are unclear are shown by curved, double-headed arrows. The capital letters represent the relations between the observed measures and their corresponding constructs; small letters represent the variances of the residuals of the observed measures. The paths between the variables are numbered.

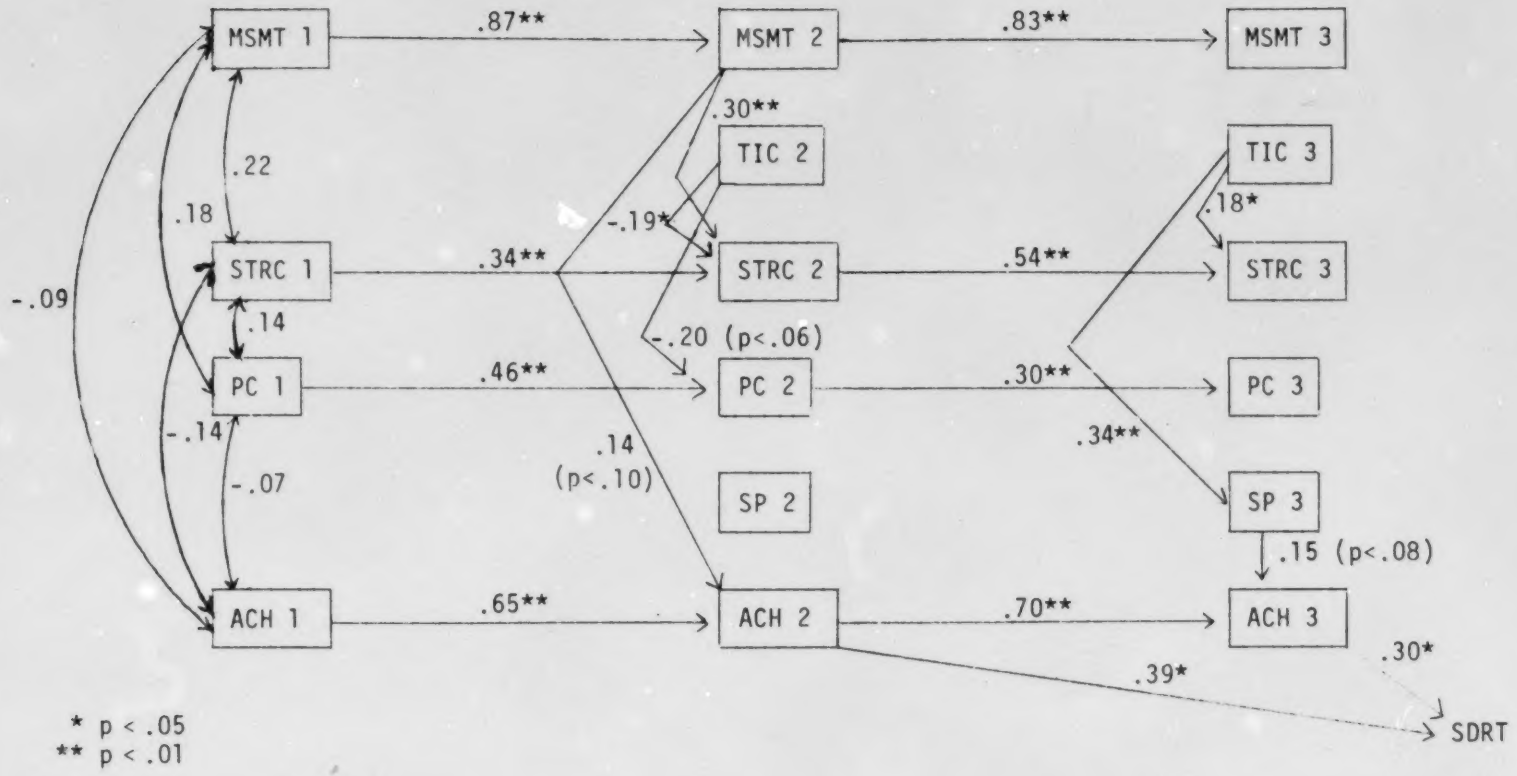

Figure 7. Path Analysis is depicting the relationships between measurement (MSMT), structure (STRC), achievement (SCH), timing of instructional change (TIC), silent practice (SP), oral practice (OP), positive consequences (PC), and the Stanford Diagnostic Reading Test (SDRT). The curved double-headed arrows represent initial relationships and the straight arrows represent causal relationships.

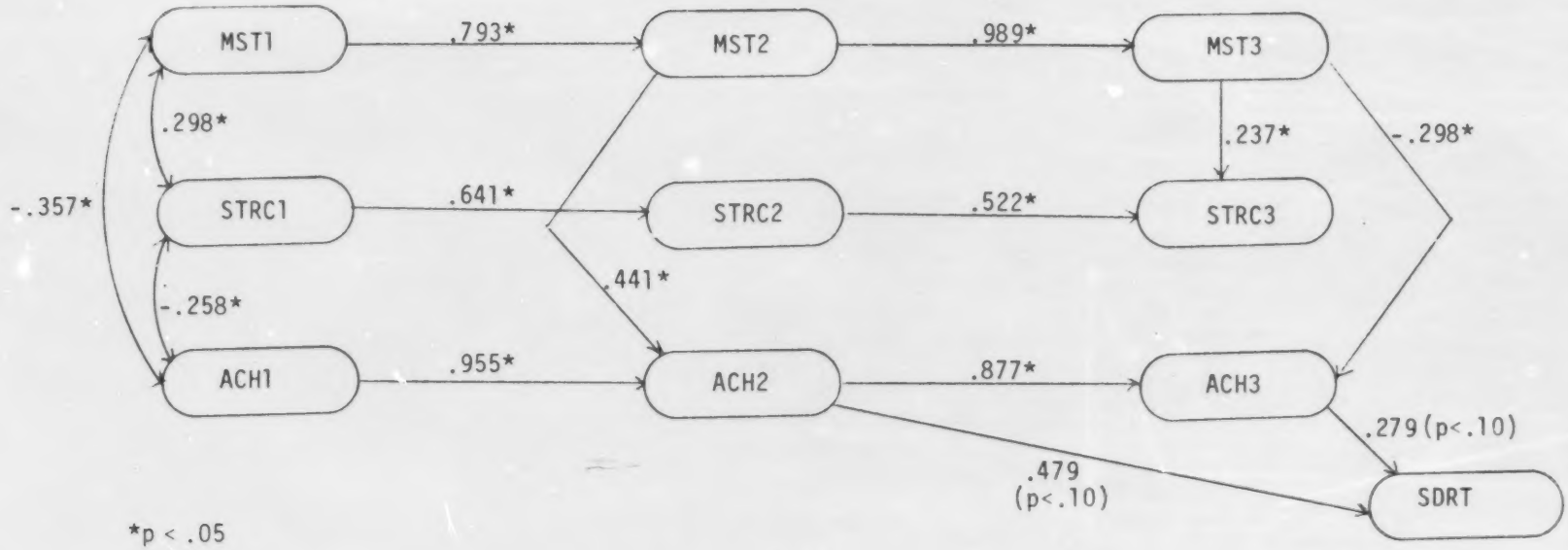

Figure 8. Structural model depicting the relationships between measurement (MST), structure (STRC), achievement (ACH), and the Stanford Diagnostic Reading Test (SDRT). The curved arrows represent initial relationships and the straight arrows indicate causal relationships.

Appendix A

Accuracy of Implementation Rating Scale

School:_____ Student:_____

Date:_____ Teacher:_____

Observer (Items 1 and 2):_____

Rater (Items 3-13):_____

Number of observations prior to rating:_____

Time observation begins:_____ Time observation ends:_____

Time allocated to reading instruction per day:_____

Curriculum used for measurement: Publisher_____

Series_____ Level_____

Instructions

Circle the number that accurately reflects your rating for each variable. Only one number may be circled per variable. 1 reflects a low level of implementation and 5 means total implementation of the Procedures to Develop and Monitor Progress on IEP Goals. See Operational Definitions. Items 1 and 2 require direct observation of the measurement administration. Items 3, 4, 5, 6, and 7 require inspection of the student graph. Items 8, 9, and 10 require inspection of the student's IEP form. The Instructional Plan must be inspected to rate item 11. The Change Record must be inspected to rate items 12 and 13.

Dir. Obs.	1. Administering the Measurement Task	1	2	3	4	5
	2. Selecting the Stimulus Material	1	2	3	4	5
Inspect Graph	3. Sampling for Instructional Level	1	2	3	4	5
	4. Baseline	1	2	3	4	5
	5. Graph Set-up	1	2	3	4	5
	6. Aimline	1	2	3	4	5
	7. Timing of Instructional Changes	1	2	3	4	5
Insp. IEP	8. Long-Range Goal	1	2	3	4	5
	9. Short-Term Objective	1	2	3	4	5
	10. Measurement System	1	2	3	4	5
Insp. Inst. Plan	11. Instructional Plan	1	2	3	4	5
Insp. Change Rec.	12. Substantial Changes	1	2	3	4	5
	13. One, Clear Change	1	2	3	4	5

AIRS

Operational Definitions

Accuracy of Implementation Rating Scale

1. Administering the Measurement Task

5 - The measurement task is administered correctly: teacher brings stopwatch and pencil to measurement area; gives correct directions for the task; administers the measurement procedure for one minute; correctly marks the teacher copy; correctly counts words correct and incorrect; correctly counts words correct and incorrect; correctly plots the data point.

1 - The teacher: forgets necessary materials; does not give directions; does not time the task accurately; fails to mark the teacher copy or incorrectly marks errors; miscounts correct and incorrect words; and inaccurately plots the data point.

2. Selecting the Stimulus Material

5 - The teacher has followed these procedures: Uses passages selected from the level that represents the annual goal. Observers should record the book from which the passage was selected and later check this with the long-range goal level. At this level find the pages in these stories that do not have excessive dialogue, indentations, and/or unusual pronouns. Write these page numbers on equal size slips of paper.

- Put the slips of paper into a drawbag and shake it.

- Randomly pick a slip of paper.

- The page number chosen is the page where the student begins reading. If the page chosen is a passage that was read earlier during the week, draw another page number.

Other completely random procedures are also rated a 5. If, however, not all passages have an equal chance of being selected, a 4 rating would be indicated.

1 - The teacher fails to randomly pick the passage or the sample is taken from a domain which is greater or smaller than the one indicated in the goal.

3. Sampling for Instructional Level

5 - The teacher has sampled from higher or lower reading levels to find the level in which the student reads 20-29 wpm (grades 1 & 2) or 30-39 wpm (grades 3 and up).

1 - The teacher is measuring at a level which is too high or too low.

4. Baseline

5 - The student's performance has been measured at least 3 times to establish a stable baseline. A stable baseline means that all data points fall within a range of 10.

1 - The teacher has not found a level for which a stable baseline has been established or has failed to collect 3 data points during the baseline phase.

5. Graph Set-Up

5 - The graph is accurately set up: The dates filled in on the horizontal axis; the vertical axis is correctly labeled words read per minute from _____ material; the units of measurement are specified; the student's name and subject area are certified; a key identifies the symbols for correct (.) and incorrect (x); symbols are placed at the intersection of date and score; the data points are connected with straight lines; and absences are recorded on the graph as (abs.).

1 - The graph does not include many of the items mentioned above.

6. Aimline

5 - The long-range goal is marked on the graph with an X at the intersection of the desired performance level and date of attainment and a line of desired progress connects the point representing the student's median score of the last 3 data points from baseline and the LRG.

1 - The long-range goal is not marked on the graph and/or the median and LRG are not connected.

7. Timing of Instructional Changes

5 - All the adjustments in the student's program are made at the appropriate time given the rules for data utilization:

(1) Compare the actual slope based on 7 to 10 data points to the slope required to attain the Annual Goal.

(2) If the actual slope is equal to, or steeper than, the Annual Goal slope, continue the program.

(3) If the actual slope is flatter than the Annual Goal slope, change the program.

1 - None of the adjustments in the student's program are made at the appropriate time.

8. Long-Range Goal

5 - The long-range goal is accurately written; goal specifies the number of weeks until next review; stimulus materials for the goal represents the level in which the student is performing at entry level criterion; goal specifies student behavior; goal specifies mastery criterion of 50 wpm with fewer than 5 errors (grades 1 & 2) or 70 wpm with fewer than 7 errors (grades 3-5) when there are 36 weeks until the annual review. If there are fewer than 36 weeks, the criteria can be lowered proportionately.

1 - The long-range goal contains none of the above criteria.

9. Short-Term Objective

5 - The short-term objective is accurately written; stimulus material and behavior is specified; and the average increase in performance is the desired performance minus the actual performance divided by the number of weeks until the annual review.

1 - The short-term objective contains none of the above criteria.

10. Measurement System

5 - The teacher has indicated how the material is organized, the frequency of measurement, and what is to be recorded on the graph.

1 - The measurement system is not specified.

11. Instructional Plan

5 - The instructional plan includes clear and specific descriptions of the instructional procedures, the time spent in each activity, the pertinent materials, the arrangements, and the motivational strategies.

1 - The instructional plan is unclear and lacks specific descriptions of the instructional procedures, the time spent in each activity, the pertinent materials, the arrangements, and the motivational strategies.

12. Substantial Changes

5 - The adjustments in the student's program are always substantial (have a good chance of being effective; see Unit XIV).

1 - The adjustments are never substantial.

13. Clear Change

5 - All the adjustments made introduce only one, clear program change.

1 - All the adjustments made introduce more than one change and/or the change is unclear.

Appendix B

Structure of Instruction Rating Scale (SIRS)

School:_____ Student:_____

Date:_____ Teacher:_____

Observer:_____ Number of Students in Group:_____

Number of observations prior to rating:_____

Time observation begins:_____ Time observation ends:_____

Time allocated to reading instruction per day:_____

Curriculum used for instruction: Publisher_____

Series_____ Level_____

Instructions

Circle the number that accurately reflects your rating for each variable. Only one number may be circled per variable. If you are unable to evaluate a certain variable, mark N/A (not applicable) next to the left-hand column.

1.	Instructional Grouping	1	2	3	4	5
2.	Teacher-directed Learning	1	2	3	4	5
3.	Active Academic Responding	1	2	3	4	5
4.	Demonstration/Prompting	1	2	3	4	5
5.	Controlled Practice	1	2	3	4	5
6.	Frequency of Correct Answers	1	2	3	4	5
7.	Independent Practice	1	2	3	4	5
8.	Corrections	1	2	3	4	5
9.	Positive Consequences	1	2	3	4	5
10.	Pacing	1	2	3	4	5
11.	Oral Practice on Outcome Behavior	1	2	3	4	5
12.	Silent Practice on Outcome Behavior	1	2	3	4	5

SIRS

Operational Definitions Codebook

1. Instructional Grouping

5 - 90% or more of the instruction this student receives from the teacher is on an individual basis.

1 - 10% or less of the instruction this student receives from the teacher is on an individual basis.

2. Teacher-Directed Learning

5 - Student's instruction is extremely organized, businesslike, and teacher is firm in direction and control of activities. For example, student is presented with questions, student has material to cover, etc.

1 - Student's instruction is casually organized and very spontaneous. Teacher is not committed to having the student work on a particular set of material. Instructional materials do not determine what activities student engages in and the lessons change according to problems or mood of this student.

3. Active Academic Responding

5 - The student is actively practicing the academic skills to be learned more than 75% of the time observed. Specifically, the student is engaged in oral or written responding to teacher questions or written material, e.g., reading aloud, answering questions, writing, or computing. Student rarely is involved in non-academic conversations with teacher or other students Attending to the lesson without responding, such as sitting, looking, listening, and/or following along in a book does not apply. The student must make an active, written or oral response.

1 - The student is actively practicing the skills to be learned less than 10% of the time observed. Instructional lessons may be interrupted or shortened to include "process" and other non-academic activities, e.g., clarifying feelings, opinions, and working on arts and crafts.

4. Demonstration and Prompting

5 - Appropriate steps of the desired behavior to be performed are demonstrated for the student. Student is given an opportunity to practice the step(s) as teacher provides prompts for correct behavior that approximates or achives desired response.

1 - Teacher attempts to teach the student a behavior without using demonstration and prompting techniques.

SIRS

5. Controlled Practice

5 - Student's practice of material is actively controlled by teacher who frequently asks questions to clarify that the student understands what has just been demonstrated. Questions are convergent (single factual answer) and the student's answers consistently follow the questions and are given teacher feedback.

1 - Student is rarely questioned by teacher following demonstration of new materials. Questions are more divergent (open-ended, several interpretations) than convergent (single factual answer). Student's response is not consistently followed by teacher feedback. The type of questions are such that several answers are acceptable, i.e., questions are abstract or ambiguous.

Examples:

If during an oral reading session:

a) the teacher frequently attempts to clarify the material with convergent questions ("what color hat was John wearing?"), a 5 would be recorded.

b) the teacher asks few questions, most of which are divergent ("What do you think this means?"), a 1 would be recorded.

c) the teacher asks few convergent questions or many divergent questions, the appropriate rating would be a 3.

6. Frequency of Correct Answers

5 - Academic lessons are conducted in such a way that the difficulty of the material allows the student to achieve mean accuracy of 80% or higher.

1 - Academic material is difficult for student, component steps are large or unsequenced, and mean accuracy for student is less than 55%.

(Note: If the student has no opportunity for oral or written response during the observational period, item 6 would be rated N/A - not applicable, while items 3 and 5 would most likely be rated 1).

7. Independent Practice

5 - When engaged in independent seatwork, the student frequently is monitored by the teacher who assists, clarifies, and praises the student for academic engaged tasks.

(Note: Independent seatwork is defined here as a student working on an assigned task for at least 5 minutes. [If no such 5-minute block of time is observed, Item 7 is rated N/A].)

SIRS

1 - When student is engaged in academic seat-work activities, little attention is given by teacher who directs seat-work activities from a distance or engages in work separate from the assigned seat work. Teacher is generally not helpful or supportive to student during independent practice time.

8. Corrections

5 - The student's errors are consistently corrected by the teacher. When the student either does not respond, responds incorrectly, or does not respond in unison if the activity is group directed and requires such responding, the teacher will systematically attempt to correct the student by asking a simpler question, refocusing student's attention to elicit correct response from the student or provide general rules by which to determine the correct answer 90% or more of the time.

1 - Student's errors are rarely and inconsistently corrected by the teacher. The student responses are not systematically corrected. Student's errors are corrected 50% or less of the time.

For example: In oral reading this includes teacher correction of skips and mispronunciations, or help in sounding out hesitations.

9. Positive Consequences

5 - Positive events (tokens, points, activities, etc.) are given to the student when performing the desired behavior. When learning a new skill the student receives positive consequence for approximations of the desired behavior. Consequences are consistently received during academic training time. Praise and compliments, e.g., "good working, nice job," are not included in this definition.

1 - Student rarely receives positive consequences for academic work. When student receives consequences they usually are for social behavior, rather than for behaviors occurring under systematic academic training.

10. Pacing

5 - The pace of the lesson is rapid, providing many opportunities for response by the student. As a result, attention is high and off-task behavior is low.

1 - The pace of the lesson is slow and the student's rate of responding is low. Lesson format frequently varies, is not highly structured, and student attention may be low.

SIRS

11. Oral Practice on Outcome Behavior

5 - Student reads aloud from context nearly all the time (85-100% or 12-15 min. of a 15 min. observation).

1 - Student does not read aloud during the observation (0% of the time).

(Note: Reading aloud for measurement purposes should not be considered when rating this variable. Reading in context is defined as reading phrases, sentences, paragraphs, or story selections.)

Examples:

If the student is reading isolated words nearly the entire time, the appropriate rating is a 3.

If the student is reading aloud from a text about half the time, a 3 would be recorded.

12. Silent Practice on Outcome Behavior

5 - Student reads silently from context nearly all the time (85-100% or 12-15 min. of a 15 min. observation).

1 - Student does not read silently during the observation (0% of the time).

(Note: Reading in context is defined as the same as #11. The examples of #11 are the same for #12, with silent reading.)

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art (Research Report No. 1). November, 1977.

Ysseldyke, J. E., & Regan, R. R. Nondiscriminatory assessment and decision making (Monograph No. 7). February, 1979.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the PIAT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota roundtable conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives (Research Report No. 7). June, 1979.

Algozzine, B. The disturbing child: A validation report (Research Report No. 8). June, 1979.

Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.

Ysseldyke, J. E., Algozzine, B., Regan, R., & Potter, M. Technical adequacy of tests used by professionals in simulated decision making (Research Report No. 9). July, 1979.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). August, 1979.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). August, 1979.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled (Research Report No. 11). August, 1979.

Deno, S. L., Chiang, B., Tindal, G., & Blackburn, M. Experimental analysis of program components: An approach to research in CSDC's (Research Report No. 12). August, 1979.

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. Similarities and differences between underachievers and students labeled learning disabled: Identical twins with different mothers (Research Report No. 13). September, 1979.

Ysseldyke, J., & Algozzine, R. Perspectives on assessment of learning disabled students (Monograph No. 11). October, 1979.

Poland, S. F., Ysseldyke, J. E., Thurlow, M. L., & Mirkin, P. K. Current assessment and decision-making practices in school settings as reported by directors of special education (Research Report No. 14). November, 1979.

McGue, M., Shinn, M., & Ysseldyke, J. Validity of the Woodcock-Johnson psycho-educational battery with learning disabled students (Research Report No. 15). November, 1979.

Deno, S., Mirkin, P., & Shinn, M. Behavioral perspectives on the assessment of learning disabled children (Monograph No. 12). November, 1979.

Sutherland, J. H., Algozzine, B., Ysseldyke, J. E., & Young, S. What can I say after I say LD? (Research Report No. 16). December, 1979.

Deno, S. L., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance (Monograph No. 13). December, 1979.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. The influence of test scores and naturally-occurring pupil characteristics on psycho-educational decision making with children (Research Report No. 17). December, 1979.

Algozzine, B., & Ysseldyke, J. E. Decision makers' prediction of students' academic difficulties as a function of referral information (Research Report No. 18). December, 1979.

Ysseldyke, J. E., & Algozzine, B. Diagnostic classification decisions as a function of referral information (Research Report No. 19). January, 1980.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). January, 1980.

Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehnle, K. Relationships among simple measures of spelling and performance on standardized achievement tests (Research Report No. 21). January, 1980.

Deno, S. L., Mirkin, P. K., & Marston, D. Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). January, 1980.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). January, 1980.

Deno, S. L., Mirkin, P. K., Robinson, S., & Evans, P. Relationships among classroom observations of social adjustment and sociometric rating scales (Research Report No. 24). January, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Factors influential on the psychoeducational decisions reached by teams of educators (Research Report No. 25). February, 1980.

Ysseldyke, J. E., & Algozzine, B. Diagnostic decision making in individuals susceptible to biasing information presented in the referral case folder (Research Report No. 26). March, 1980.

Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.

Ysseldyke, J. E., Regan, R. R., & Schwartz, S. Z. The use of technically adequate tests in psychoeducational decision making (Research Report No. 28). April, 1980.

Richey, L., Potter, M., & Ysseldyke, J. Teachers' expectations for the siblings of learning disabled and non-learning disabled students: A pilot study (Research Report No. 29). May, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Instructional planning: Information collected by school psychologists vs. information considered useful by teachers (Research Report No. 30). June, 1980.

Algozzine, B., Webber, J., Campbell, M., Moore, S., & Gilliam, J. Classroom decision making as a function of diagnostic labels and perceived competence (Research Report No. 31). June, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., Richey, L., & Thurlow, M. L. Psychoeducational assessment and decision making: A computer-simulated investigation (Research Report No. 32). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., & Richey, L. Psychoeducational assessment and decision making: Individual case studies (Research Report No. 33). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R., Potter, M., & Richey, L. Technical supplement for computer-simulated investigations of the psychoeducational assessment and decision-making process (Research Report No. 34). July, 1980.

Algozzine, B., Stevens, L., Costello, C., Beattie, J., & Schmid, R. Classroom perspectives of LD and other special education teachers (Research Report No. 35). July, 1980.

Algozzine, B., Siders, J., Siders, J., & Beattie, J. Using assessment information to plan reading instructional programs: Error analysis and word attack skills (Monograph No. 14). July, 1980.

Ysseldyke, J., Shinn, M., & Epps, S. A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability (Research Report No. 36). July, 1980.

Algozzine, B., & Ysseldyke, J. E. An analysis of difference score reliabilities on three measures with a sample of low achieving youngsters (Research Report No. 37). August, 1980.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 38). August, 1980.

Richey, L. S., Ysseldyke, J., Potter, M., Regan, R. R., & Greener, J. Teachers' attitudes and expectations for siblings of learning disabled children (Research Report No. 39). August, 1980.

Ysseldyke, J. E., Algozzine, B., & Thurlow, M. L. (Eds.). A naturalistic investigation of special education team meetings (Research Report No. 40). August, 1980.

Meyers, B., Meyers, J., & Deno, S. Formative evaluation and teacher decision making: A follow-up investigation (Research Report No. 41). September, 1980.

Fuchs, D., Garwick, D. R., Featherstone, N., & Fuchs, L. S. On the determinants and prediction of handicapped children's differential test performance with familiar and unfamiliar examiners (Research Report No. 42). September, 1980.

Algozzine, B., & Stoller, L. Effects of labels and competence on teachers' attributions for a student (Research Report No. 43). September, 1980.

Ysseldyke, J. E., & Thurlow, M. L. (Eds.). The special education assessment and decision-making process: Seven case studies (Research Report No. 44). September, 1980.

Ysseldyke, J. E., Algozzine, B., Potter, M., & Regan, R. A descriptive study of students enrolled in a program for the severely learning disabled (Research Report No. 45). September, 1980.

Marston, D. Analysis of subtest scatter on the tests of cognitive ability from the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 46). October, 1980.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. Identifying children with learning disabilities: When is a discrepancy severe? (Research Report No. 47). November, 1980.

Fuchs, L., Tindal, J., & Deno, S. Effects of varying item domain and sample duration on technical characteristics of daily measures in reading (Research Report No. 48). January, 1981.

Marston, D., Lowry, L., Deno, S., & Mirkin, P. An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study (Research Report No. 49). January, 1981.

Marston, D., & Deno, S. The reliability of simple, direct measures of written expression (Research Report No. 50). January, 1981.

Epps, S., McGue, M., & Ysseldyke, J. E. Inter-judge agreement in classifying students as learning disabled (Research Report No. 51). February, 1981.

Epps, S., Ysseldyke, J. E., & McGue, M. Differentiating LD and non-LD students: "I know one when I see one" (Research Report No. 52). March, 1981.

Evans, P. R., & Peham, M. A. S. Testing and measurement in occupational therapy. A review of current practice with special emphasis on the Southern California Sensory Integration Tests (Monograph No. 15). April, 1981.

Fuchs, L., Wesson, C., Tindal, G., & Mirkin, P. Teacher efficiency in continuous evaluation of IEP goals (Research Report No. 53). June, 1981.

Fuchs, D., Featherstone, N., Garwick, D. R., & Fuchs, L. S. The importance of situational factors and task demands to handicapped children's test performance (Research Report No. 54). June, 1981.

Tindal, G., & Deno, S. L. Daily measurement of reading: Effects of varying the size of the item pool (Research Report No. 55). July, 1981.

Fuchs, L. S., & Deno, S. L. A comparison of teacher judgment, standardized tests, and curriculum-based approaches to reading placement (Research Report No. 56). August, 1981.

Fuchs, L., & Deno, S. The relationship between curriculum-based mastery measures and standardized achievement tests in reading (Research Report No. 57). August, 1981.

Christenson, S., Graden, J., Potter, M., & Ysseldyke, J. Current research on psychoeducational assessment and decision making: Implications for training and practice (Monograph No. 16). September, 1981.

Christenson, S., Ysseldyke, J., & Algozzine, B. Institutional constraints and external pressures influencing referral decisions (Research Report No. 58). October, 1981.

Fuchs, L., Fuchs, D., & Deno, S. Reliability and validity of curriculum-based informal reading inventories (Research Report No. 59). October, 1981.

Algozzine, B., Christenson, S., & Ysseldyke, J. Probabilities associated with the referral-to-placement process (Research Report No. 60). November, 1981.

Tindal, G., Fuchs, L. Christenson, S., Mirkin, P., & Deno, S. The relationship between student achievement and teacher assessment of short- or long-term goals (Research Report No. 61). November, 1981.

Mirkin, P., Fuchs, L., Tindal, G., Christenson, S., & Deno, S. The effect of IEP monitoring strategies on teacher behavior (Research Report No. 62). December, 1981.

Wesson, C., Mirkin, P., & Deno, S. Teachers' use of self instructional materials for learning procedures for developing and monitoring progress on IEP goals (Research Report No. 63). January, 1982.

Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). January, 1982.

Potter, M., & Mirkin, P. Instructional planning and implementation practices of elementary and secondary resource room teachers: Is there a difference? (Research Report No. 65). January, 1982.

Thurlow, M. L., & Ysseldyke, J. E. Teachers' beliefs about LD students (Research Report No. 66). January, 1982.

Graden, J., Thurlow, M. L., & Ysseldyke, J. E. Academic engaged time and its relationship to learning: A review of the literature (Monograph No. 17). January, 1982.

King, R., Wesson, C., & Deno, S. Direct and frequent measurement of student performance: Does it take too much time? (Research Report No. 67). February, 1982.

Greener, J. W., & Thurlow, M. L. Teacher opinions about professional education training programs (Research Report No. 68). March, 1982.

Algozzine, B., & Ysseldyke, J. Learning disabilities as a subset of school failure: The oversophistication of a concept (Research Report No. 69). March, 1982.

Fuchs, D., Zern, D. S., & Fuchs, L. S. A microanalysis of participant behavior in familiar and unfamiliar test conditions (Research Report No. 70). March, 1982.

Shinn, M. R., Ysseldyke, J., Deno, S., & Tindal, G. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving (Research Report No. 71). March, 1982.

Thurlow, M. L. Graden, J., Greener, J. W., & Ysseldyke, J. E. Academic responding time for LD and non-LD students (Research Report No. 72). April, 1982.

Graden, J., Thurlow, M., & Ysseldyke, J. Instructional ecology and academic responding time for students at three levels of teacher-perceived behavioral competence (Research Report No. 73). April, 1982.

Algozzine, B., Ysseldyke, J., & Christenson, S. The influence of teachers' tolerances for specific kinds of behaviors on their ratings of a third grade student (Research Report No. 74). April, 1982.

Wesson, C., Deno, S., & Mirkin, P. Research on developing and monitoring progress on IEP goals: Current findings and implications for practice (Monograph No. 18). April, 1982.

Mirkin, P., Marston, D., & Deno, S. L. Direct and repeated measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students (Research Report No. 75). May, 1982.

Algozzine, B., Ysseldyke, J., Christenson, S., & Thurlow, M. Teachers' intervention choices for children exhibiting different behaviors in school (Research Report No. 76). June, 1982.

Tucker, J., Stevens, L. J., & Ysseldyke, J. E. Learning disabilities: The experts speak out (Research Report No. 77). June, 1982.

Thurlow, M. L., Ysseldyke, J. E., Graden, J., Greener, J. W., & Mecklenberg, C. Academic responding time for LD students receiving different levels of special education services (Research Report No. 78). June, 1982.

Graden, J. L., Thurlow, M. L., Ysseldyke, J. E., & Algozzine, B. Instructional ecology and academic responding time for students in different reading groups (Research Report No. 79). July, 1982.

Mirkin, P. K., & Potter, M. L. A survey of program planning and implementation practices of LD teachers (Research Report No. 80). July, 1982.

Fuchs, L. S., Fuchs, D., & Warren, L. M. Special education practice in evaluating student progress toward goals (Research Report No. 81). July, 1982.

Kuehnle, K., Deno, S. L., & Mirkin, P. K. Behavioral measurement of social adjustment: What behaviors? What setting? (Research Report No. 82). July, 1982.

Fuchs, D., Dailey, Ann Madsen, & Fuchs, L. S. Examiner familiarity and the relation between qualitative and quantitative indices of expressive language (Research Report No. 83). July, 1982.

Videen, J., Deno, S., & Marston, D. Correct word sequences: A valid indicator of proficiency in written expression (Research Report No. 84). July, 1982.

Potter, M. L. Application of a decision theory model to eligibility and classification decisions in special education (Research Report No. 85). July, 1982.

Greener, J. E., Thurlow, M. L., Graden, J. L., & Ysseldyke, J. E. The educational environment and students' responding times as a function of students' teacher-perceived academic competence (Research Report No. 86). August, 1982.

Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study (Research Report No. 87). August, 1982.

Skiba, R., Wesson, C., & Deno, S. L. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 88). September, 1982.

Marston, D., Tindal, G., & Deno, S. L. Eligibility for learning disability services: A direct and repeated measurement approach (Research Report No. 89). September, 1982.

Thurlow, M. L., Ysseldyke, J. E., & Graden, J. L. LD students' active academic responding in regular and resource classrooms (Research Report No. 90). September, 1982.

Ysseldyke, J. E., Christenson, S., Pianta, R., Thurlow, M. L., & Algozzine, B. An analysis of current practice in referring students for psychoeducational evaluation: Implications for change (Research Report No. 91). October, 1982.

Ysseldyke, J. E., Algozzine, B., & Epps, S. A logical and empirical analysis of current practices in classifying students as handicapped (Research Report No. 92). October, 1982.

Tindal, G., Marston, D., Deno, S. L., & Germann, G. Curriculum differences in direct repeated measures of reading (Research Report No. 93). October, 1982.

Fuchs, L.S., Deno, S. L., & Marston, D. Use of aggregation to improve the reliability of simple direct measures of academic performance (Research Report No. 94). October, 1982.

Ysseldyke, J. E., Thurlow, M. L., Mecklenburg, C., & Graden, J. Observed changes in instruction and student responding as a function of referral and special education placement (Research Report No. 95). October, 1982.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Effects of frequent curriculum-based measurement and evaluation on student achievement and knowledge of performance: An experimental study (Research Report No. 96). November, 1982.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Direct and frequent measurement and evaluation: Effects on instruction and estimates of student progress (Research Report No. 97). November, 1982.

Tindal, G., Wesson, C., Germann, G., Deno, S. L., & Mirkin, P. K. The Pine County model for special education delivery: A data-based system (Monograph No. 19). November, 1982.

Epps, S., Ysseldyke, J. E., & Algozzine, B. An analysis of the conceptual framework underlying definitions of learning disabilities (Research Report No. 98). November, 1982.

Epps, S., Ysseldyke, J. E., & Algozzine, B. Public-policy implications of different definitions of learning disabilities (Research Report No. 99). November, 1982.

Ysseldyke, J. E., Thurlow, M. L., Graden, J. L., Wesson, C., Deno, S. L., & Algozzine, B. Generalizations from five years of research on assessment and decision making (Research Report No. 100). November, 1982.

Marston, D., & Deno, S. L. Measuring academic progress of students with learning difficulties: A comparison of the semi-logarithmic chart and equal interval graph paper (Research Report No. 101). November, 1982.

Beattie, S., Grise, P., & Algozzine, B. Effects of test modifications on minimum competency test performance of third grade learning disabled students (Research Report No. 102). December, 1982

Algozzine, B., Ysseldyke, J. E., & Christenson, S. An analysis of the incidence of special class placement: The masses are burgeoning (Research Report No. 103). December, 1982.

Marston, D., Tindal, G., & Deno, S. L. Predictive efficiency of direct, repeated measurement: An analysis of cost and accuracy in classification (Research Report No. 104). December, 1982.

Wesson, C., Deno, S., Mirkin, P., Sevcik, B., Skiba, R., King, R., Tindal, G., & Maruyama, G. Teaching structure and student achievement effects of curriculum-based measurement: A causal (structural) analysis (Research Report No. 105). December, 1982.

Mirkin, P. K., Fuchs, L. S., & Deno, S. L. (Eds.). Considerations for designing a continuous evaluation system: An integrative review (Monograph No. 20). December, 1982.